ltcc3.tex Week 3 15 October 2015

Chapter III. MARTINGALES.

§1. DISCRETE-PARAMETER MARTINGALES

We summarise what we need; for details, see Williams[**W**], or Neveu [**Nev**].

Definition. A process $X = (X_n)$ is called a *martingale* relative to $((\mathcal{F}_n), P)$ if
(i) X is adapted (to $(\mathcal{F}_n)$),
(ii) $E|X_n| < \infty$ for all n,
(iii) $E[X_n|\mathcal{F}_{n-1}] = X_{n-1} \qquad P-a.s. \qquad (n \geq 1)$;
X is a *supermartingale* if in place of (iii)

$$E[X_n|\mathcal{F}_{n-1}] \leq X_{n-1} \qquad P-a.s. \qquad (n \geq 1);$$

X is a *submartingale* if in place of (iii)

$$E[X_n|\mathcal{F}_{n-1}] \geq X_{n-1} \qquad P-a.s. \qquad (n \geq 1).$$

Thus: a martingale is 'constant on average', and models a *fair* game;
a supermartingale is 'decreasing on average', and models an *unfavourable game*;
a submartingale is 'increasing on average', and models a *favourable* game.

Note. 1. Martingales have many connections with harmonic functions in probabilistic potential theory. The terminology in the inequalities above comes from this: supermartingales correspond to superharmonic functions, submartingales to subharmonic functions.
2. X is a submartingale [supermartingale] iff $-X$ is a supermartingale [submartingale];
X is a martingale iff it is both a submartingale and a supermartingale .
3. (X_n) is a martingale iff $(X_n - X_0)$ is a martingale . So we may without loss of generality take $X_0 = 0$ when convenient.
4. If X is a martingale , then for $m < n$

$$E[X_n|\mathcal{F}_m] = E[E(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m] \qquad \text{(iterated conditional expectations)}$$

$$= E[X_{n-1}|\mathcal{F}_m] \qquad a.s. \qquad \text{(martingale property)}$$

$$= \cdots = E[X_m|\mathcal{F}_m] \qquad a.s. \qquad \text{(induction on } n\text{)},$$

$$= X_m \qquad (X_m \text{ is } \mathcal{F}_m\text{-measurable})$$

and similarly for submartingales, supermartingales.
5. Examples of a martingale include: sums of independent, integrable zero-mean random variables [submartingale : positive mean; supermartingale : negative mean].

From the OED: martingale (etymology unknown)

1. 1589. An article of harness, to control a horse's head.
2. Naut. A rope for guying down the jib-boom to the dolphin-striker.
3. A system of gambling which consists in doubling the stake when losing in order to recoup oneself (1815).

Thackeray: 'You have not played as yet? Do not do so; above all avoid a martingale if you do.'

Problem. Analyse this strategy.

Gambling games have been studied since time immemorial – indeed, the Pascal-Fermat correspondence of 1654 which started the subject was on a problem (de Méré's problem) related to gambling.

The doubling strategy above has been known at least since 1815.

The term 'martingale ' in our sense is due to J. VILLE (1939). Martingales were studied by Paul LÉVY (1886-1971) from 1934 on [see obituary, *Annals of Probability* **1** (1973), 5-6] and by J. L. DOOB (1910-2004) from 1940 on. The first systematic exposition was Doob [D], Ch. VII.

Example: Accumulating data about a random variable ([**W**], 96, 166-167).
If $\xi \in L_1(\Omega, \mathcal{F}, P)$, $M_n := E(\xi|\mathcal{F}_n)$ (so M_n represents our best estimate of ξ based on knowledge at time n), then by iterated conditional expectations

$$E[M_n|\mathcal{F}_{n-1}] = E[E(\xi|\mathcal{F}_n)|\mathcal{F}_{n-1}] = E[\xi|\mathcal{F}_{n-1}] = M_{n-1},$$

so (M_n) is a martingale . One has the convergence

$$M_n \to M_\infty := E[\xi|\mathcal{F}_\infty] \qquad a.s. \qquad \text{and in } L_1;$$

see below.

§2. MARTINGALE CONVERGENCE.

A supermartingale is 'decreasing on average'. Recall that a decreasing sequence [of real numbers] that is bounded below converges (decreases to its greatest lower bound or infimum). This suggests that a supermartingale which is bounded below converges a.s. This is so [Doob's Forward Convergence Theorem: [**W**], §§11.5, 11.7].

More is true. Call X L_1-bounded if

$$\sup_n E|X_n| < \infty.$$

THEOREM (DOOB). *An L_1-bounded supermartingale is a.s. convergent: there exists X_∞ finite such that*

$$X_n \to X_\infty \qquad (n \to \infty) \qquad a.s.$$

In particular, we have

DOOB'S MARTINGALE CONVERGENCE THEOREM [W, §11.5]. *An L_1-bounded martingale converges a.s.*

We say that

$$X_n \to X_\infty \text{ in } L_1$$

if

$$E|X_n - X_\infty| \to 0 \qquad (n \to \infty).$$

For a class of martingales, one gets convergence in L_1 as well as almost surely [= with probability one].

Theorem ([N, IV.2], [W, Ch.14]). *The following are equivalent for martingales $X = (X_n)$:*
(i) X_n converges in L_1,
(ii) X_n is L_1-bounded, and its a.s. limit X_∞ (which exists, by above) satisfies

$$X_n = E[X_\infty | \mathcal{F}_n],$$

(iii) There exists an integrable random variable X with

$$X_n = E[X | \mathcal{F}_n].$$

Such martingales are called *regular* [**N**] or *uniformly integrable* (UI) [**W**], or *closed*; the limit X_∞ is said to *close* the martingale.

§3. MARTINGALE TRANSFORMS.

Now think of a gambling game, or series of speculative investments, in discrete time. There is no play at time 0; there are plays at times $n = 1, 2, \cdots$, and

$$\Delta X_n := X_n - X_{n-1}$$

represents our net winnings per unit stake at play n. Thus if X_n is a martingale, the game is 'fair on average'.

Call a process $C = (C_n)_{n=1}^{\infty}$ *previsible* (or *predictable*) if

$$C_n \text{ is } \mathcal{F}_{n-1} - \text{measurable for all } n \geq 1.$$

Think of C_n as your stake on play n (C_0 is not defined, as there is no play at time 0). Previsibility says that you have to decide how much to stake on play n based on the history *before* time n (i.e., up to and including play $n-1$). Your winnings on game n are $C_n \Delta X_n = C_n(X_n - X_{n-1})$. Your total (net) winnings up to time n are

$$Y_n = \Sigma_{k=1}^n C_k \Delta X_k = \Sigma_{k=1}^n C_k (X_k - X_{k-1}).$$

We write

$$Y = C \bullet X, \qquad Y_n = (C \bullet X)_n, \qquad \Delta Y_n = C_n \Delta X_n$$

($(C \bullet X)_0 = 0$ as Σ_1^0 is empty), and call $C \bullet X$ the *martingale transform* of X by C.

Theorem. *(i) If C is a bounded non-negative previsible process and X is a supermartingale, $C \bullet X$ is a supermartingale null at zero.*
(ii) If C is bounded and previsible and X is a martingale, $C \bullet X$ is a martingale null at zero.

Proof. With $Y = C \bullet X$ as above,

$$E[Y_n - Y_{n-1} | \mathcal{F}_{n-1}] = E[C_n (X_n - X_{n-1}) | \mathcal{F}_{n-1}]$$

$$= C_n E[(X_n - X_{n-1}) | \mathcal{F}_{n-1}]$$

(as C_n is bounded, so integrable, and $\mathcal{F}_{n-1}$-measurable, so can be taken out)

$$\leq 0$$

in case (i), as $C \geq 0$ and X is a supermartingale,

$$= 0$$

in case (ii), as X is a martingale. •

Interpretation. You can't beat the system!
In the martingale case, previsibility of C means we can't foresee the future (which is realistic and fair). So we expect to gain nothing – as we should.
Note. 1. Martingale transforms were introduced and studied by D. L. BURKHOLDER in 1966 [*Ann. Math. Statist.* **37**, 1494-1504]. For a textbook account, see e.g. [**N**], VIII.4.
2. Martingale transforms are the discrete analogues of stochastic integrals. They dominate the mathematical theory of finance in discrete time, just as stochastic integrals dominate the theory in continuous time. We use them to rebalance our portfolio as new price information comes in.

PROPOSITION (Martingale Transform Lemma). *An adapted sequence of real integrable random variables* (M_n) *is a martingale iff for any bounded previsible sequence* (H_n),

$$E(\Sigma_{r=1}^{n} H_r \Delta M_r) = 0 \qquad (n = 1, 2, \cdots).$$

Proof. If (M_n) is a martingale, X defined by $X_0 = 0$, $X_n = \Sigma_1^n H_r \Delta M_r \qquad (n \geq 1)$ is the martingale transform $H \bullet M$, so is a martingale.

Conversely, if the condition of the Proposition holds, choose j, and for any $\mathcal{F}_j$-measurable set A write $H_n = 0$ for $n \neq j+1$, $H_{j+1} = I_A$. Then (H_n) is previsible, so the condition of the Proposition, $E(\Sigma_1^n H_r \Delta M_r) = 0$, becomes

$$E[I_A(M_{j+1} - M_j)] = 0.$$

Since this holds for every set $A \in \mathcal{F}_j$, the definition of conditional expectation gives

$$E(M_{j+1} | \mathcal{F}_j) = M_j.$$

Since this holds for every j, (M_n) is a martingale. •

§4. STOPPING TIMES AND OPTIONAL STOPPING.

A random variable T taking values in $\{0, 1, 2, \cdots; +\infty\}$ is called a *stopping time* (or *optional time*) if

$$\{T \leq n\} = \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n \qquad \forall n \leq \infty.$$

Equivalently,

$$\{T = n\} \in \mathcal{F}_n \qquad n \leq \infty.$$

Think of T as a time at which you decide to quit a gambling game: whether or not you quit at time n depends only on the history up to and including time n – NOT the future.

The following important classical theorem is discussed in [**W**], §10.10.

THEOREM (DOOB'S OPTIONAL STOPPING THEOREM, OST). *Let* T *be a stopping time,* $X = (X_n)$ *be a supermartingale, and assume that one of the following holds:*

(i) T *is bounded [*$T(\omega) \leq K$ *for some constant* K *and all* $\omega \in \Omega$*];*

(ii) $X = (X_n)$ *is bounded [*$|X_n(\omega)| \leq K$ *for some* K *and all* n, ω*];*

(iii) $ET < \infty$ and $(X_n - X_{n-1})$ *is bounded.*

Then X_T *is integrable, and*

$$EX_T \leq EX_0.$$

If here X is a martingale, then

$$EX_T = EX_0.$$

The OST is important in many areas, such as sequential analysis in statistics, and to American options in finance (options that can be exercised at any time up to and including expiry).

Write $X_n^T := X_{n\wedge T}$ for the sequence (X_n) *stopped* at time T.

Proposition. *(i) If (X_n) is adapted and T is a stopping-time, the stopped sequence $(X_{n\wedge T})$ is adapted.*
(ii) If (X_n) is a martingale [supermartingale] and T is a stopping time, (X_n^T) is a martingale [supermartingale].

Proof. If $\phi_j := I\{j \leq T\}$,

$$X_{T\wedge n} = X_0 + \Sigma_1^n \phi_j (X_j - X_{j-1}).$$

Since $\{j \leq T\}$ is the complement of $\{T < j\} = \{T \leq j-1\} \in \mathcal{F}_{j-1}$, $\phi_j = I\{j \leq T\} \in \mathcal{F}_{j-1}$, so (ϕ_n) is previsible. So (X_n^T) is adapted.

If (X_n) is a martingale, so is (X_n^T) as it is the martingale transform of (X_n) by (ϕ_n). Since by previsibility of (ϕ_n)

$$E(X_{T\wedge n}|\mathcal{F}_{n-1}) = X_0 + \Sigma_1^{n-1} \phi_j (X_j - X_{j-1}) + \phi_n (E[X_n|\mathcal{F}_{n-1}] - X_{n-1}),$$

$\phi_n \geq 0$ shows that if (X_n) is a supermartingale [submartingale], so is $(X_{T\wedge n})$. •

§5. DOOB DECOMPOSITION.

THEOREM. *Let $X = (X_n)$ be an adapted process with each $X_n \in L_1$. Then X has an (essentially unique) Doob decomposition*

$$X = X_0 + M + A : \qquad X_n = X_0 + M_n + A_n \qquad \forall n \qquad (D)$$

with M a martingale null at zero, A a previsible process null at zero. If also X is a submartingale ('increasing on average'), A is increasing: $A_n \leq A_{n+1}$ for all n, a.s.

Proof. If X has a Doob decomposition (D),

$$E[X_n - X_{n-1}|\mathcal{F}_{n-1}] = E[M_n - M_{n-1}|\mathcal{F}_{n-1}] + E[A_n - A_{n-1}|\mathcal{F}_{n-1}].$$

The first term on the right is zero, as M is a martingale. The second is $A_n - A_{n-1}$, since A_n (and A_{n-1}) is $\mathcal{F}_{n-1}$-measurable by previsibility. So

$$E[X_n - X_{n-1}|\mathcal{F}_{n-1}] = A_n - A_{n-1}, \tag{1}$$

and summation gives

$$A_n = \Sigma_1^n E[X_k - X_{k-1}|\mathcal{F}_{k-1}], \qquad a.s.$$

We use this formula to *define* (A_n), clearly previsible. We then use (D) to *define* (M_n), then a martingale, giving the Doob decomposition (D).

If X is a submartingale, the LHS of (1) is ≥ 0, so the RHS of (1) is ≥ 0, i.e. (A_n) is increasing. •

Note. 1. Although the Doob decomposition is a simple result in discrete time, the analogue in continuous time is deep (see below). This illustrates the contrasts that may arise between the theories of stochastic processes in discrete and continuous time.
2. Previsible processes are 'easy' (trading is easy if you can foresee price movements!). So the Doob Decomposition splits any (adapted) process X into two bits, the 'easy' (previsible) bit A and the 'hard' (martingale) bit M. Moral: martingales are everywhere!
3. Submartingales model favourable games, so are *increasing on average*. It 'ought' to be possible to split such a process into an *increasing* bit, and a remaining 'trendless' bit. It is – the trendless bit is the martingale.

§**6. EXAMPLES.**

1. *Simple random walk.* Recall the simple random walk: $S_n := \Sigma_1^n X_k$, where the X_n are independent tosses of a fair coin, taking values ± 1 with equal probability 1/2. Suppose we decide to bet until our net gain is first $+1$, then quit. Let T be the time we quit; T is a stopping time. It has been analysed in detail; see e.g. Grimmett & Stirzaker [**GS**], §5.2. From this, note:
(i) $T < \infty$ a.s.: the gambler will certainly achieve a net gain of $+1$ eventually;
(ii) $ET = +\infty$: the mean waiting-time till this happens is infinity.
Hence also:
(iii) No bound can be imposed on the gambler's maximum net loss before his net gain first becomes $+1$.

At first sight, this looks like a foolproof way to make money out of nothing: just bet till you get ahead (which happens eventually, by (i)), then quit. However, as a gambling strategy, this is hopelessly impractical: because of (ii), you need unlimited time, and because of (iii), you need unlimited capital - neither of which is realistic.

Notice that the Optional Stopping Theorem fails here: we start at zero, so $S_0 = 0$, $ES_0 = 0$; but $S_T = 1$, so $ES_T = 1$. This example shows two things:
a) The Optional Stopping Theorem does indeed need conditions, as the conclusion may fail otherwise [none of the conditions (i) – (iii) in the OST are satisfied in the example above],
(b) Any practical gambling (or trading) strategy needs to have some integrability or boundedness restrictions to eliminate such theoretically possible but practically ridiculous cases.
2. *The doubling strategy.* The strategy of doubling when losing – *the* martingale, according to the Oxford English Dictionary (§3) has similar properties – and would be suicidal in practice as a result.
3. *The Saint Petersburg Game.* A single play of the Saint Petersburg game consists of a sequence of coin tosses stopped at the first head; if this is the rth toss, the player receives a prize of \$ 2^r. [Thus the expected gain is $\Sigma_1^\infty 2^{-r}.2^r = +\infty$, so the random variable is not integrable, and martingale theory does not apply.] Let S_n denote the player's cumulative gain after n plays of the game. The question arises as to what the 'fair price' of a ticket to play the game is. It turns out that fair prices exist (in a suitable sense), but the fair price of the nth play varies with n – surprising, as all the plays are the replicas of each other.

Other examples may be constructed of games which are 'fair' in some sense, but in which the player sustains a net loss, tending to $-\infty$, with probability one. For a discussion of such examples, see e.g. Feller [F1], X.3,4.

§7. CONTINUOUS-PARAMETER MARTINGALES

The martingale property in continuous time is just that suggested by the discrete-time case:

$$E[X_t|\mathcal{F}_s] = X_s \qquad (s < t),$$

and similarly for submartingales and supermartingales. There are regularization results, under which one can take X_t right-continuous in t. Among the contrasts with the discrete case, we mention that the Doob-Meyer decomposition, easy in discrete time, is a deep result in continuous time. For background, see e.g. Meyer [Mey66], and subsequent work by Meyer and the French school (Dellacherie & Meyer, *Probabilités et potentiel*, I-V, etc.).

§8. POISSON PROCESSES; LÉVY PROCESSES

Suppose we have a process $X = (X_t : t \geq 0)$ which has *stationary independent increments*: if $X_{t+u} - X_t$ denotes the increment over the interval $(t, t+u]$, then
(i) the distribution of the increments depends only on the length u of the interval, not on

its starting-point t (stationarity);
(ii) increments over disjoint intervals are independent.
Such a process is called a *Lévy process*, in honour of their creator, the great French probabilist Paul Lévy (1886-1971) [see *Ann. Probab.* **1**.1 for his obituary, by Loève]. Then for each $n = 1, 2, \ldots,$

$$X_t = X_{t/n} + (X_{2t/n} - X_{t/n}) + \ldots + (X_t - X_{(n-1)t/n})$$

displays X_t as the sum of n independent (by independent increments), identically distributed (by stationary increments) random variables. Consequently, X_t is *infinitely divisible*: for each n, it is the sum of n independent identically distributed random variables. The characteristic functions (CFs) of infinitely divisible distributions are known, and are given by the Lévy-Khintchine formula (L-K); see e.g. Bertoin [Ber]. The prime example is (anticipating Week 4):

the Wiener process, or Brownian motion, is a Lévy process.

Poisson Processes.

The increment $N_{t+u} - N_u$ $(t, u \geq 0)$ of a Poisson process is the number of failures in $(u, t+u]$ (in the language of renewal theory – see Week 2). By the lack-of-memory property of the exponential, this is independent of the failures in $[0, u]$, so the increments of N are *independent.* It is also identically distributed to the number of failures in $[0, t]$, so the increments of N are *stationary.* That is, N has stationary independent increments, so is a Lévy process:

Poisson processes are Lévy processes.

We need an important property: two Poisson processes (on the same filtration) are independent iff they never jump together (a.s.). For proof, see e.g. Revuz & Yor [R-Y], XII.1.

The Poisson count in an interval of length t is Poisson $P(\lambda t)$ (where the rate λ is the parameter in the exponential $E(\lambda)$ of the renewal-theory viewpoint), and the Poisson counts of disjoint intervals are independent. This extends from intervals to Borel sets:
(i) For a Borel set B, the Poisson count in B is Poisson $P(\lambda|B|)$, where $|.|$ denotes Lebesgue measure;
(ii) Poisson counts over disjoint Borel sets are independent.

Poisson (Random) Measures.

If ν is a finite measure, call a random measure ϕ *Poisson* with *intensity* (or characteristic) *measure* ν if for each Borel set B, $\phi(B)$ has a Poisson distribution with parameter $\nu(B)$, and for $B_1, \ldots, B_n$, $\phi(B_1), \ldots, \phi(B_n)$ are independent. One can extend to σ-finite measures ν: if (E_n) are disjoint with union $\mathbb{R}$ and each $\nu(E_n) < \infty$, construct ϕ_n from ν

restricted to E_n and write ϕ for $\sum \phi_n$.

Poisson Point Processes.

With ν as above a (σ-finite) measure on $\mathbb{R}$, consider the product measure $\mu = \nu \times dt$ on $\mathbb{R} \times [0, \infty)$, and a Poisson measure ϕ on it with intensity μ. Then ϕ has the form

$$\phi = \sum_{t \geq 0} \delta_{(e(t),t)},$$

where the sum is *countable* (for background and details, see [Ber], §0.5, whose treatment we follow here). Thus ϕ is the sum of Dirac measures over 'Poisson points' $e(t)$ occurring at Poisson times t. Call $e = (e(t) : t \geq 0)$ a *Poisson point process* with *characteristic measure* ν,

$$e = Ppp(\nu).$$

For each Borel set B,

$$N(t, B) := \phi(B \times [0, t]) = card\{s \leq t : e(s) \in B\}$$

is the *counting process* of B – it counts the Poisson points in B – and is a Poisson process with rate (parameter) $\nu(B)$. All this reverses: starting with an $e = (e(t) : t \geq 0)$ whose counting processes over Borel sets B are Poisson $P(\nu(B))$, then – as no point can contribute to more than one count over disjoint sets – disjoint counting processes never jump together, so are independent by above, and $\phi := \sum_{t \geq 0} \delta_{(e(t),t)}$ is a Poisson measure with intensity $\mu = \nu \times dt$.

Note. The link between point processes and martingales goes back to S. Watanabe in 1964 (Japanese J. Math.). The approach via Poisson point processes is due to Kiyosi Itô (1915-2008) in 1970 (Proc. 6th Berkeley Symp.); see below, and – in the context of excursion theory – [R-W2], VI §8. For a monograph treatment of Poisson processes, see [Kin].

Compound Poisson processes.

A random variable Poisson distributed with parameter λ has generating function $\sum_{n=0}^{\infty} e^{-\lambda} \lambda^n / n! . s^n = \exp\{-\lambda(1-s)\}$ and CF $\exp\{-\lambda(1-e^{it})\}$. A Poisson process $Ppp(\lambda)$ jumps by 1 at Poisson points distributed with intensity λ. Now suppose that at the nth Poisson point there is a jump of size X_n, where the X_n are independent and identically distributed (iid) random variables with distribution function F and CF $\phi(t)$. The resulting process $X = (X_t)$ is called a *compound Poisson process* with *intensity* λ and *jump law* F. As above, X_t has CF $\exp\{-\lambda(1 - \phi(t))\}$. In a sense made precise by the Lévy-Khintchine formula and the Lévy-Itô decomposition, a general Lévy process may be built up from a deterministic 'drift' ct, a Brownian motion (Week 4) and a limit of sums of compound Poisson processes, 'compensated' by having their means subtracted (these compensated sums are then martingales). For details, see e.g. Bertoin [Ber].